In the claims:

1. A method of speaker-dependent voice command recognition comprising the steps of:

providing a hybrid of sentence network and Gaussian mixture models with a shared pool of distributions; and

performing an out-of-vocabulary procedure based on the score difference between a top candidate and background model over the recognized in-vocabulary word.

2. The method of Claim 1 wherein said network is a three section network to represent speech embedded in extra speech where first and last sections are intended to absorb extra- speech and the middle section to match with in-vocabulary speech.
3. The method of Claim 2 wherein the first and last sections of the network comprise fully interconnected nodes and the second section comprises nodes sequentially (left –to-right) connected.

4. The method of Claim 3 wherein to each of the nodes is attached a power density function (PDF) and the PDF attached to the first and last sections are shared by the nodes belonging to the two sections.

5. The method of Claim 4 wherein the PDFs in the network are modeled as single Gaussian distributions with a unique variance shared by all nodes of the network

6. The method of Claim 5 wherein the PDFs of the second section are trained from the enrollment utterances of a given command.
7. The method of Claim 6 wherein the PDF of the first and last sections are the centroids of a clustering of the mean vectors of the PDFs of the second section.
8. The method of Claim 7 wherein transition from one node to another is attached to a weight, and the balance between recognition errors of utterance with or without extra speech is controlled by adjusting the weights of the components of the nodes of the first and last sections.
9. The method of Claim 8 wherein an utterance is accepted as containing in-vocabulary word based on a rejection parameter, which has several alternative forms.
10. The method of Claim 9 wherein the rejection parameter is calculated using the following steps:

    calculating, the best possible log-likelihood using a three section network model,

    locating the first and last frame of the in-vocabulary word,

    extracting the cumulate log likelihood from the first to the last frame of the in-vocabulary word,

    calculating the best possible log-likelihood using a network model representing only the extra-speech from the first to the last frame of the in-vocabulary word and

    dividing the difference of the above two values of log likelihood by the

number of frames of the in-vocabulary word.

11. The method of Claim 9 wherein the rejection parameter is calculated by the following steps:

calculating a, the best possible log-likelihood using a three section network model;

locating the first and last frame of the in-vocabulary word;

calculating the best possible log-likelihood using a network model representing only the extra-speech for three sections:

-- from beginning of the utterance to first frame of the in-vocabulary word,

-- from the first to last frame of the in-vocabulary word,

-- from the last frame of the in-vocabulary word to the end of utterance;

subtracting from a the above three values; and

dividing the resulting value by number of frames of the in-vocabulary word.

12. The method of Claim 9 wherein the rejection parameter is calculated by the following steps:

calculating the best possible log-likelihood using a three section network model;

locating the first and last frame of the in-vocabulary word;

calculating the best possible log-likelihood using a network model representing only the extra-speech over the whole utterance; and

dividing the difference of the above two values of log likelihood by the number of frames of the in-vocabulary word.